ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.20, No.10, October 2009, pp.2787−2798
doi: 10.3724/SP.J.1001.2009.03373

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

无线传感器网络中基于协作水印的虚假数据识别和过滤算法*

易叶青[1], 林亚平[1,2+], 李小龙[1], 羊四清[1], 尤志强[2]

[1](湖南大学 计算机与通信学院,长沙 410082)

[2](湖南大学 软件学院,长沙 410082)

A False Data Recognition and Filtering Algorithm Using Cooperation Watermarks for Wireless Sensor Networks

YI Ye-Qing[1], LIN Ya-Ping[1,2+], LI Xiao-Long[1], YANG Si-Qing[1], YOU Zhi-Qiang[2]

[1](School of Computer and Communication, Hunan University, Changsha 410082, China)

[2](School of Software, Hunan University, Changsha 410082, China)

+ Corresponding author: E-mail: yplin@hnu.cn

Yi YQ, Lin YP, Li XL, Yang SW, You ZQ. A false data recognition and filtering algorithm using cooperation watermarks for wireless sensor networks. *Journal of Software*, 2009,20(10):2787−2798. http://www.jos.org.cn/1000-9825/3373.htm

Abstract: In this paper, a data authenticate algorithm based on cooperation watermarks is proposed for recognition and filtering false data and replayed packets, in which two kinds of watermarks are embed into data packets:One is robust watermark for the authentication of sender's identification and the freshness of data, the other is Semi-Fragile watermark for the authentication of content, generated by *t* witnesses. There are several good features in the proposed algorithm. Firstly, different watermarks have no interaction one another. Secondly, single sensor node can independently extract the watermark to validate data packets while nodes have no ability to modify and fabricate watermark. Theoretically and experimentally, the algorithm has good performance of the peak value signal to noise ratio and signal to noise ratio by embedding watermarks into packets under most circumstances. And the algorithm is of high sensibility to malicious modified data, of robustness to some certain noise disturbance, lossy compression and so on. The new algorithm has lower communication cost and higher capability for the false data recognition and filtering when comparing with existing algorithm based on MAC(Message Authentication Code) schemes.

Key words: wireless sensor network; cooperation watermarks; false data recognition and filtering

摘 要: 本文提出一种基于协作水印的数据认证算法来识别虚假数据和重复包,算法在每个数据包中嵌入两类水印:一类是鲁棒性水印,用于对发送者的身份和数据的新鲜性进行认证;另一类是由 *t* 个证人节点协作生成、嵌入的

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z227 (国家高技术研究发展计划(863)); the Natural Science Foundation of Hu'nan Province of China under Grant No.06JJ20049 (湖南省自然科学基金); the Scientific Research Fund of Hu'nan Provincial Education Department of China under Grant No.06B047 (湖南省教育厅优秀青年项目); the (湖南省重点建设学科项目)

Received 2007-09-20; Accepted 2008-04-15

半脆弱水印,用于对数据内容进行认证.算法保证了多个水印之间互不影响;算法允许网络中的单个节点独立地提取水印,验证数据包的正确性,却不能伪造或修改水印.仿真和分析表明算法在数据包中嵌入水印后大多数情况下均有较好的峰值信噪比和信噪比;同时算法能对恶意篡改数据具有较高的敏感性,对一定程度噪声干扰、有损压缩等具有较好的鲁棒性.算法与已有的基于 MAC(Message Authentication Code)的虚假数据过滤算法相比具有更低的通信开销和更高的识别与过滤虚假数据的能力.

关键词: 无线传感器网络;协作水印;虚假数据识别与过滤

中图法分类号: TP393 **文献标识码**: A

随着通信技术、嵌入式计算技术和传感器技术的飞速发展和日益成熟,无线传感器网络(wireless sensor networks)引起了人们极大关注[1,2].传感器网络的主要目的是收集传感数据,如何保证收集到的数据具有新鲜性、完整性、正确性是至关重要的[3].

然而,传感器网络节点通常部署在野外或者是敌方区域,攻击者可以通过俘获节点向网络灌注虚假数据、恶意篡改正在传送的数据包、发送过时数据包,若不加防范,这些虚假的数据将会引发错误警报,干扰用户决策,消耗有限的网络资源,造成严重的后果[4].鉴于虚假数据的安全威胁,不少学者提出了一些对付虚假数据的办法[5−14],它们的共同特点是在待发送的数据包后额外附加 t 个 MAC(message authentication code),并在数据转发的过程中对数据包进行认证,从而实现对虚假数据的识别和过滤.这些已有的工作存在以下问题:暴露在传输数据外面的 MAC 容易造成安全隐患;传感器网络节点通过无线信道通信,数据在传输过程中容易遭到噪声干扰,尽管数据受到了噪声的干扰,当附加在数据包后的 MAC 信息受到噪声干扰后,上述方案会将这些数据包误认为是虚假数据包,从而造成可用数据无法传输到基站的问题;传感器网络内部也会对数据进行一些有损操作,这同样会导致附加在数据包后的 MAC 受损,从而造成正确的数据无法传送到基站的问题;此外,这种方法大多数难以抵抗重放攻击.

为此,本文提出了一种基于协作水印的虚假数据识别和过滤算法,算法针对传感器网络特征,给出一种适合传感器网络的多节点协作的数字水印认证算法,该算法以时间戳和簇头节点的身份信息为鲁棒性水印,以证人节点认证信息为半脆弱水印,然后将这多个水印以协同方式嵌入到聚合后的数据中,从而达到对发送节点身份与发送数据内容双重认证的目的.本文所提出的算法采用多个证人节点协作生成、嵌入水印信息,算法允许网络中的单个节点独立的提取水印验证数据的合法性、正确性、完整性,这种技术一方面能在数据转发过程中尽可能早地过滤虚假数据包,节约有限的网络资源,另一方面尽管允许网络中的单个节点独立的提取水印,却限制了任何单个节点无法伪造或修改水印,从而保证了数据的安全性;算法不仅能有效地过滤虚假数据包或被恶意篡改的数据包,而且对一定程度噪声干扰、有损压缩等具有较好的鲁棒性.

本文第 1 节概述相关工作.第 2 节提出多节点协作水印的虚假数据的识别和过滤算法.第 3 节对算法过滤虚假数据的能力、开销等进行理论上的定性分析.第 4 节进行相应的仿真实验.第 5 节对本文的工作进行总结.

1 相关工作

Ye 等人率先对传感器网络中的虚假数据识别和过滤问题进行了讨论,提出了 SEF 虚假数据过滤机制[4,5],其基本思想是在发送的数据包后附加 t 个 MAC 来实施认证,由于每个数据包后附加了 t 个 MAC,攻击者要想伪造虚假数据就必须获得 t 个不同组的密钥,从而增加了攻击难度.另外算法允许中间节点以一定概率对数据进行认证,并对虚假数据进行过滤,达到节约网络资源的目的.

为了提高安全性和虚假数据的过滤概率,一些学者在 Fan Ye 等人的基础上提出了相应的改进方案.Zhu 等人在 SEF 算法的基础上,提出了在路由时进行交叉的逐步认证机制[6],该机制能保证虚假数据在一定跳数内被发现并且被过滤.Yang 等人认为上述方法都是基于对称密钥机制.其安全性不够,提出了一种基于密码交换(commutative cipher)的路由过滤机制[7].该机制假设各节点与基站共享一会话密钥(session key),路由中报文转发节点不知道报文源节点的会话密钥,但可通过交换密码对报文进行检测,从而提高了安全性.Yang 等人提出利

用部署节点的先验知识和二元多项式构建认证密钥的方法来识别和过滤虚假数据[8].Ma 等人认为上述方法受门限值的限制,因此提出一种不受门限值 t 限制的虚假数据过滤机制[9],该机制假设存在一种比普通节点能量强得多的节点作为簇头完成数据聚合,聚合结果必须附加簇内各个普通节点产生的 MAC 才算合法,目的节点在收到汇聚结果时通过对聚合结果及附加 MAC 的校验,实现对虚假数据的过滤.Li 等人提出能将俘获节点定位的虚假数据过滤机制[10],并采用单向哈希链技术,保证中间节点只能校验数据但不能修改数据.本实验室成员胡玉鹏,林亚平等针对无线传感器网络通信环境不可靠,易受虚假数据的攻击的特点,提出一种基于动态哈希树的鲁棒虚假数据过滤方案,能有效提升虚假数据过滤能力和鲁棒性[11].此外,针对不同网络拓扑和路由机制,一些学者在文献[4,5]的基础上提出了适应于网络拓扑和路由机制的虚假数据过滤策略[12−14].

数字水印技术是安全研究领域的一个分支,具有难于提取、难于修改、难于伪造的特性,在数据认证、防伪等领域中得以应用[15],Feng 等人提出利用传感器节点定位时允许在距离的期望值与测量值之间有一定误差,并通过修改测量值嵌入水印的方法[16],该方法利用传感器节点感知数据可能有微量误差的特性,只要嵌入水印时引入的误差在限定范围内,不会影响感知数据的正常使用;以上方法表明在无线传感器网络中采用数字水印技术来保护传感器网络中的感知数据是一种切实可行的方法.但上述方法对于虚假数据过滤和识别问题仍是不适合的.

2 多节点协作水印的虚假数据识别和过滤算法

传感器网络节点一般可通过飞行器投放,因此假定同一簇的节点是一起投放的,它们的物理位置彼此靠近,节点分布较为密集,大部分节点能直接通信,故在同一个簇内节点所采集的数据往往具有较高时空相关性[17];本文假定基站是在用户的直接控制之下或是被安置在一个安全的地方,不可能被攻击者所俘获[5−14];另外,假设簇头节点的确定是根据节点的能量状况采用轮换机制[18].

本节,提出一种适合传感器网络的基于多节点协作数字水印(multiple node cooperation watermarks,简称 MNCW)的虚假数据识别和过滤算法.

2.1 基本思想

算法的基本思想:首先为算法建立一个较大的密钥池,但只有基站有权知道整个密钥池的密钥及其编号,每个传感器节点在投放之前按一定的策略从该密钥池中选取一个较小的密钥环(key ring)预置在该节点内;当簇头接收到基站的请求数据命令后,再向簇内每个节点发出收集数据的命令,每个节点接收到簇头的命令后,将自身采集的数据发送给簇头;簇头对接收到的全部数据进行聚合操作,得到的数据记为 D;然后将数据 D 进行数学变换(如:小波变换)$D\rightarrow\{D_1,D_2,D_3,\ldots\}$,不妨令 D_1 为数据 D 的最重要部分(如:小波变换的低频系数),再将簇头的身份信息和当前的时间戳以鲁棒性水印的方式嵌入到 D_1 得 D'_1;簇头节点为数据 D'_1 寻找包括簇头在内的 t 个证人节点 $\{node_1,node_2,...,node_t\}$,并由证人节点生成 t 个水印信息发送给簇头,簇头将这 t 个水印协作嵌到 $\{D_2,D_3,\ldots\}$中得$\{D'_2,D'_3,\ldots\}$,再将$\{D'_1,D'_2,D'_3,\ldots\}$进行逆变换得 D',最后簇头节点将数据 D'和 t 个用于产生和嵌入水印的密钥编号组成一个数据包,通过中转节点以多跳的形式发送给基站,数据包在传输的过程中,中间节点可以以一定的概率来验证数据包的正确性,当数据传输到基站时,基站再对数据包进行全面验证,从而全面剔除虚假数据包.

2.2 支持多节点协作水印的密钥预置算法

尽管目前已有不少文献研究无线传感器网络的密钥管理和预置,但一般都是以加密数据为目的,其追求的是邻居节点具有较高的密钥共享概率,因此难以支持我们所提出的多节点协作水印技术.与已有的密钥管理方法不同的是,我们所需要的密钥预置策略一方面要保证除 sink 节点之外其他单个节点难以伪造协作水印;另一方面也要保证任意两个节点存在一定概率共享一定数量的密钥.我们解决该问题的思路是:在传感器网络部署前,为其构造一个密钥池 $\Omega=\{\lambda_1,\lambda_2,...,\lambda_n\}$,其中 λ_i 为密钥子集,各密钥子集的大小相等且满足: $\lambda_i\cap\lambda_j=\varnothing$.令每密钥子集 λ_i 的密钥个数为 m, $\lambda_i=\{k_j\mid i\times m\le j\le(i+1)\times m-1\}$,其中 k_j 为密钥,显然密钥池的总密钥数为 nm.在本

文的方案中每个密钥 k_j 包括两个部分,即: $k_j=\langle k_{jE},k_{jd}\rangle$,其中 k_{jE} 用于加密水印信息,k_{jd} 用水印信息的嵌入和提取,并要求 $0<k_{jd}<1$.系统再为所有的密钥进行统一编号,编号由密钥所属子集号和子集内序号两部分组成,密钥 k_j 的编号记为:id_{kj}.在每个节点部署之前,用户为该节点随机地选择一个子密钥集λ_i,并从该密钥集λ_i 中随机选取 $\theta(\theta<m)$个密钥和相应的密钥编号预置到该节点.

源节点在生成嵌入水印时需要 t 个不同组的证人节点密钥信息,因此达到除 Sink 节点之外其他单个节点难以伪造协作水印,而同一组的节点又存在一定概率共享密钥的目的.

2.3 多节点协作水印技术

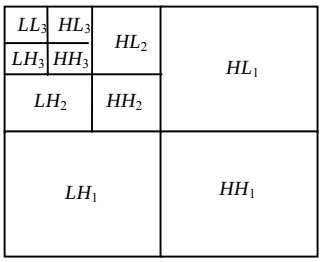

Fig.1 Coefficients of three level wavelet decomposition for data D

图 1 数据 D 的三级小波变换系数

当簇头对收集到的数据进行聚合操作后得到聚合结果为 D,不妨令 D 大小为 $row\times col$ 的二维矩阵,数据 D 按如下的方法添加水印.

步骤 2.3.1. 嵌入用于判断是否为重复包的鲁棒性水印.

簇头首先对数据 D 进行三级 Haar 二维小波变换(本文的方法不局限于小波变换,也可以为其他的数学变换,如:离散余弦变换等)得到如图 1 所示的小波系数 $\{LL_3,LH_3,HL_3,HH_3,LH_2,HL_2,HH_2,LH_1,HL_1,HH_1\}$,其中 LL_3,表示数据 D 的三级小波变换的低频系数,代表数据 D 的最重要信息.然后,簇头将自身的身份信息(如:节点的 ID 号)和当前的时间戳进行编码,并转换成二进制序列 $WR=[wr_1,wr_2,...,wr_L]$,其中 $wr_i\in\{0,1\}$,再将二进制序列 WR 作为鲁棒性水印,按算法 1[19]嵌入到 LL_3 中,得到 $LL_3^{'}$.

算法 1.

(1) 首先按 $\mathrm{E}(LL_3)=\frac{1}{\xi}\sum_{i=1}^{col/8}\sum_{j=1}^{row/8}LL_3(i,j)$ 计算 $\mathrm{E}(LL_3)$的值,它表示低频系数的伪均值,式中ξ用于控制水印嵌入的强度,当ξ大时表示水印嵌入强度小,当ξ小时表示水印嵌入强度大.

(2) 按下式计算 $\hbar$, $\hbar$ 代表水印的能量:

$$\hbar=\begin{cases}-\sigma+0.5q, Q(LL_3(i,j),q)=s_i\\ -\sigma+1.5q, Q(LL_3(i,j),q)\neq s_i and\ \sigma>0.5q\\ -\sigma-0.5q, Q(LL_3(i,j),q)\neq s_i and\ \sigma\le 0.5q\end{cases}.$$

式中 q 为量化步长,$0.5q$ 和 $1.5q$ 是用来把伪均值移到量化间隔的中间,使校正后的伪均值相对难以溢出现有的量化间隔中,σ为量化噪声, $\sigma=\left\lfloor\frac{\mathrm{E}(LL_3)}{q}\right\rfloor\cdot q$,$\lfloor\ \rfloor$是取下整操作,式中 $Q(LL_3(i,j),q)=\left\lfloor\frac{LL_3(i,j)}{q}\right\rfloor\%2$,其中%表示求模运算.

(3) 计算 $LL_3^{'}(i,j)=LL_3(i,j)+\hbar$.

(4) 重复(2)(3)步得 $LL_3^{'}$.

步骤 2.3.2. 为聚合数据 D 选取 t 个“证人”节点,并由“证人”节点生成 t 个用于数据认证的水印.

首先簇头收集 1 跳内邻居节点的密钥子集编号,簇头再随机选取包含簇头在内的 t 个密钥子集编号两两相异的节点,并令这些节点为 $\{node_1,node_2,...,node_t\}$ (假设 $node_1$ 为簇头),然后簇头把数据 $LL_3^{'}$ 广播给 $\{node_2,node_3,...,node_t\}$.节点 $\{node_1,node_2,...,node_t\}$ 得到数据 $LL_3^{'}$ 后,将采集到的数据和 $LL_3^{'}$ 进行相关性检测,相关性检测的目的是判断聚合数据的真实性,若证人节点采集的数据和簇头的聚合数据相关则认为聚合数据为真实的,否则认为不可靠.如果相关性超过阈值δ,则将从自身的密钥环中随机选择密钥 k_j,并利用单向函数[4]计算: $w_i=f(LL_3^{'},k_{jE}),1\le i\le t$,其中 k_{jE} 为密钥 k_j 的第 1 部分,w_i 为二进制串,并视为水印信息,然后再将水印 w_i、密钥 k_j 的编号 id_{kj} 和密钥 k_j 的第 2 部分 k_{jd} 发送给簇头节点;如果相关性低于阈值δ,则发送一个拒绝的消息.若簇头收到来自这 t 个节点所生成的水印 $\{w_1,w_2,...,w_t\}$、生成水印的密钥编号 $\{id_{k1},id_{k2},...,id_{kt}\}$ 和密钥第 2 部分

$\{k_{1d},k_{2d},...,k_{td}\}$,则按下面的方法嵌入水印;若簇头只收到 $r(r<t)$个节点成功返回消息,则需要按上述的方法再选取 $t-r$ 个节点(注意:同样要保证 t 个证人节点的密钥子集号两两相异),直到成功收集了 t 节点的成功返回消息.若簇头选取了∇(∇为系统参数)次仍无法找到满足上述条件的 t 个节点,则令簇头停止发送数据,并删除已收集和处理的数据,算法结束.

步骤 2.3.3. 水印的嵌入和数据的转发.

簇头按下述方法将 t 个水印嵌入到小波系数 $\{LH_3,HL_3,HH_3,LH_2,HL_2,HH_2,...\}$ 中并得到 $\{LH_3^{'},HL_3^{'},HH_3',LH_2',.\ HL_2',HH_2',...\}$ 水印嵌入之后簇头将删除收集到的水印、密钥编号和密钥的第 2 部分.

在这一步中,提出了一种新颖的适合于传感器网络的多水印协作嵌入算法.首先选择待嵌入的小波系数,并对其进行降维处理 $\{LH_3,HL_3,HH_3,LH_2,HL_2,HH_2,...\}\to Y$,不妨令长度为 M 的向量 $Y=[y_1,y_2,...,y_M]$ 表示降维后的小波系数,假定 t 个节点生成的水印 $w_i=[w_{i1},w_{i2},...,w_{iN}]$ $(1\le i\le t,N$ 表示每位证人节点产生的水印位数),然后再将向量 Y 划分成 N 个子向量,每个子向量的长度为 $P=\left\lfloor\frac{M}{N}\right\rfloor$,其中第 i 个子向量用 Y_i 标记.算法将在每个子向量中嵌入 t 个水印位 $w_{ji},1\le j\le t$,算法将水印嵌入问题转化成一个优化问题,下面将给出在子向量 Y_i 中嵌入 t 个水印位的数学模型.$Y_i^{'}=(1+\mu)Y_i$,其中μ按下面的数学模型计算:

目标函数: $\min\left(\left\|Y_i-Y_i'\right\|_2\right)$

约束条件:

$$\begin{cases} round\left(\dfrac{\|Y_i'\|_2}{k_{1d}}\right)\%2=w_{1i} \\ round\left(\dfrac{\|Y_i'\|_2}{k_{2d}}\right)\%2=w_{2i} \\ \quad\vdots\qquad\qquad\vdots \\ round\left(\dfrac{\|Y_i'\|_2}{k_{td}}\right)\%2=w_{ti} \end{cases}.$$

式中的目标函数 $\min\left(\left\|Y_i-Y_i'\right\|_2\right)$ 表明水印嵌入之后尽可能小地影响聚合结果 D,$\|Y_i\|_2$ 和 $\|Y_i'\|_2$ 分别表示向量 Y_i 与 Y_i' 的 2-范数,函数 $round()$表示取四舍五入.上述模型实际上是一个单变量优化问题,当前已有多种方法来求解,本文不再复述.上述的多水印嵌入方法是利用多个节点所提供的密钥协作嵌入的,这种嵌入算法能保证 t 个水印之间互不影响,同时当中间节点拥有用于加密水印和生成水印的 t 个密钥$\{k_1,k_2,...,k_t\}$中的任意一个密钥 k_i 时,能独立通过提取水印来判断数据包的正确性,但不能篡改水印及其他信息.

簇头用上述的数学模型求出 $\mu_i(1\le i\le t)$ 并计算出 $Y_i'(1\le i\le t)$,由数据 $Y_i'(1\le i\le t)$ 得到修改后的小波系数(即含水印的小波系数) $\{LH_3',HL_3',HH_3',LH_2',HL_2',HH_2',...\}$,簇头再将修改后的小波系数进行逆小波变换得到数据 D',簇头将密钥编号和数据 D'组成一个数据包以多跳的方式向基站进行传送.

2.4 虚假数据的剔除

当一个节点接收到一个数据包,节点首先检查是否附加了 t 个密钥编号以及编号是否合法,再通过附加的密钥编号判断这 t 个密钥是否来自 t 个不同的密钥子集,如果数据包中附加的密钥编号数不等于 t 或者这 t 个密钥不是来自 t 个不同的密钥子集则节点将该数据包视为虚假数据包,并进行删除. 如果节点满足上述条件则调用算法 2 来处理数据包.

算法2.

$[LL_3',LH_3',HL_3',HH_3',LH_2',HL_2',HH_2',LH_1',HL_1',HH_1']=DWT3(D')$;// $DWT3(*)$ 为三级小波变换函数

k=1

for i=1 to $col/8$

for j=1 to $row/8$

$\{S[k]=\left\lfloor \frac{LL_3'(i,j)}{q} \right\rfloor \%2$;k=k+1;//其中q为一常量,表示量化步长$\}$

$S \to [ID\_CusterHead, time]$;//将S按第2.3节中的步骤2.3.1的方法逆转换为簇头的身份信息和时间戳;

根据 $ID\_CusterHead$,$time$ 和节点保存的通信记录来判断是否为重复包;

If(数据包为重复包) 删除数据包,停止发送.算法结束;

else {

If 中转节点共享一个用于生成水印的密钥 $k_c=\langle k_{cE}, k_{cd}\rangle$ //通过附加的密钥编号来判定

{ $\{LH_3', HL_3', HH_3', LH_2', HL_2', HH_2', ...\} \to Y'$;//选取待嵌入的小波系数按第1.3节的方法进行降维;

将向量划分成长度为 $P=\left\lfloor \frac{M}{N} \right\rfloor$ 的N个子向量,即: $Y'=[Y_1', Y_2', ..., Y_N']$;

for i=1 to N

$w[i]=round\left(\frac{\|Y_i'\|_2}{k_{cd}}\right)\%2$; //中间节点独立提取水印

计算 $w^*=f(LL_3', k_{cE})$;//函数f(*)和第1.3节的单向函数相同

计算相关性 $R=\dfrac{\sum_{i=1}^{N}(2\cdot w[i]-1)\cdot(2\cdot w^*[i]-1)}{\sqrt{\sum_{i=1}^{N}(2\cdot w[i]-1)^2\cdot\sum_{i=1}^{N}(2\cdot w^*[i]-1)^2}}$

$=\dfrac{1}{N}\sum_{i=1}^{N}(2\cdot w[i]-1)\cdot(2\cdot w^*[i]-1).$

If $R<\delta$ //δ 为阈值

将数据包视为虚假数据,删除.算法结束.

else 将数据包发送到下一个节点.算法结束.

}

else 将数据包发送到下一个节点.算法结束.

}

上述算法对数据包进行两个方面的验证,首先检验数据包是否为重复包,其基本思想是读出嵌入在低频系数中的鲁棒水印,这些水印信息是产生数据包的簇头节点的身份信息和时间戳,因此中转节点可以利用以往的通信记录来判断是否为重复包,由于在嵌入这种鲁棒水印时没有对其进行加密处理,所以任何节点都可以进行检验,由于有多个协作的水印保护低频数据,因此尽管每个节点都可以提取鲁棒水印,却不能伪造或恶意篡改水印信息.其次算法要验证当前的数据包是否是伪造的数据,或者是否遭受恶意的篡改.当中转节点共享一个用于生成 t 个水印的密钥时,通过提取嵌入在高频系数中的水印信息来分析数据的正确性.当中转节点没有共享这种密钥时,节点无法验证数据的正确性,也就是说上述算法并不能将所有的虚假数据过滤掉,仍存在虚假数据逃逸这种过滤机制的可能.但随着中转跳数的增加,虚假数据被过滤的概率将会急剧增加,也就是说单个节点的过滤能力是有限的,但整个中转路由的节点协作过滤能力将是较强的.

2.5 基站校验

当基站接收到数据包时,因为基站有权知道密钥池的所有密钥,故能验证所有水印的完整性,任何已逃逸中转节点过滤的虚假数据或被恶意篡改的数据包将被检测到.在接收到数据包后,基站同样要检测数据包是否附加了 t 个密钥编号,再通过附加的密钥编号判断这 t 个密钥是否来自 t 个不同的密钥子集,如果数据包中附加的密钥编号数不等于 t 或者这 t 个密钥不是来自 t 个不同的密钥子集,则基站将该数据包视为虚假数据包,并进行删除.如果满足上述条件,需要判断该数据包是否为重复数据包,判断的方法和第 2.4 节的方法相同,如果是重复包则将该包丢弃,否则按照第 2.4 节的方法提取高频系数中的 t 个水印,然后逐一计算 t 个水印和通过单向函数

$f(LL_3', k_{iE})$, $1 \le i \le t$ 计算的原始水印进行相关性分析,若存在一个水印的相关性小于阈值δ,则将该数据包作虚假数据包处理,并删除.

3 算法分析

算法依赖于 t 个证人节点(它们所预置的密钥来自 t 个不同的密钥子集)所产生、协作嵌入的水印信息来验证数据包是否为虚假数据包,因此攻击者如果已经获得 t 个以上的不同密钥子集的密钥,则能伪造出算法不能有效识别的虚假数据包.下文中,我们讨论在攻击者只获得 T_c $(0 \le T_c \le t-1)$ 个不同密钥子集的密钥情况下,算法对虚假数据的过滤能力.另外由于算法在数据的低频系数中嵌入了发送者的身份信息和时间戳,当攻击者企图将重复包发送给它的邻居,邻居节点可以通过提取的鲁棒性水印和通信记录来判断该数据包是否为重复包.

由上述对算法的描述可知,当攻击者只获得 T_c 个不同密钥子集的密钥时,很难伪造含有正确水印的数据包,为了能让其他节点接受虚假数据包攻击者必须伪造 $t–T_c$ 个不同密钥子集的合法密钥编号和 $t–T_c$ 个水印,因此中转节点只要拥有 $t–T_c$ 个伪造的密钥就能过滤该虚假数据包.若攻击者不伪造密钥和水印或者伪造非法的密钥编号,则其他节点将不会接收该数据包.通过上面的分析不难得出如下的结论:

定理 1. 若密钥池分为 n 个密钥子集,每个密钥子集的密钥数为 m,每个节点内预置的密钥数为θ,攻击者获得的密钥数为 T_c,则若经过 h 跳后虚假数据被过滤的概率为 $p_h = 1-\left(1-\frac{\theta(t-T_c)}{\Sigma}\right)^h$ (Σ为密钥池密钥总数).

证明:由于每个节点只能在一个密钥子集中选取密钥,由定理 1 可知任意节点恰好选中这 $t–T_c$ 个密钥中的一个的概率为 $\frac{\theta(t-T_c)}{\Sigma}$,故没有选中这$t–T_c$个密钥中的一个密钥的概率为$1-\frac{\theta(t-T_c)}{\Sigma}$,任意$h$个节点没有选中这 $t–T_c$ 个密钥中的一个密钥的概率为$\left(1-\frac{\theta(t-T_c)}{\Sigma}\right)^h$.显然,任意 h 个节点选中这 $t–T_c$ 个密钥中的一个密钥的概率为$1-\left(1-\frac{\theta(t-T_c)}{\Sigma}\right)^h$.同理可得:经过 h 跳后虚假数据被过滤的概率为 $p_h = 1-\left(1-\frac{\theta(t-T_c)}{\Sigma}\right)^h$. □

推论 1. 当参数 t 增大时,同等情况下算法对虚假数据的过滤概率将增大.

推论 2. 当$\frac{\theta}{\Sigma}$增大时,同等情况下,算法对虚假数据的过滤概率将增大.

推论 1 说明了随着参数 t 的增大算法对虚假数据的过滤能将有所提高,但并不是参数 t 可以无限制增大,因为参数 t 增大选取 t 个证人的难度会变大,对原始聚合数据的破坏也会增大(下文中有分析);推论 2 也说明当$\frac{\theta}{\Sigma}$增大时,算法对虚假数据的过滤概率将增大,同样并不是参数$\frac{\theta}{\Sigma}$可以无限制的增大,因为参数$\frac{\theta}{\Sigma}$增大意味着总密钥池的密钥数减少或者是每个节点预置的密钥数增加,如果密钥池的密钥数太少或者每个节点预置的密钥太多,攻击者只要俘获少量节点就能知道密钥池的大部分密钥,这不利于安全,同时也增大存储开销.

定理 2. 若密钥池分为 n 个密钥子集,每个密钥子集的密钥数为 m,每个节点内预置的密钥数为θ,攻击者获得的密钥数为 T_c,一个虚假数据包在网络中平均传输 $E(hops)=\frac{mn}{\theta(t-T_c)}$跳后将被过滤.

证明:由定理 1 可得
$$E(hops)=\lim_{k\to\infty}\left(\sum_i^k i\cdot\left(1-\frac{\theta(t-T_c)}{mn}\right)^{i-1}\cdot\frac{\theta(t-T_c)}{mn}\right)$$
$$=\lim_{k\to\infty}\left(\frac{\theta(t-T_c)}{mn}\cdot\sum_i^k i\cdot\left(1-\frac{\theta(t-T_c)}{mn}\right)^{i-1}\right)$$
$$=\frac{mn}{\theta(t-T_c)}.$$ □

由一阶无线模型(first order radio model)[18]可知,假定簇头向 sink 节点发送数据包的总数为 Q 个,其中重复

包占的百分比为α%,虚假数据包占的百分比为β%,经 H 跳传送到 sink 节点,若不采用本文的算法,在不改变路由的情况下通信开销为 $E_{\text{communication}}=Q\left(\sum_{i=1}^{H}\varepsilon_{amp}\times L_D\times d_i^2+H\times E_{elec}\times L_D\right)$,式中 L_D 表示数据包的 Bit 数,d_i 表示每跳的距离,式中 $\varepsilon_{amp}=100pJ/bit/m^2$ 表示发送数据的能耗参数,E_{elec}=50nJ/bit 表示接受数据的能耗参数.

假定簇头向 sink 节点发送数据包的总数为 Q 个,其中重复包占的百分比为α%,虚假数据包占的百分比为β%,经 H 跳传送到 sink 节点,若采用本文的算法,不改变路由的情况下通信开销为:

(1) 合法包共有 $Q(1-\alpha\%-\beta\%)$,这部分包能传输到 sink 节点,其耗能为

$$Q(1-\alpha\%-\beta\%)\left(\sum_{i=1}^{H}\varepsilon_{amp}\times(L_D+L_{id\_k})\times d_i^2+H\times E_{elec}\times(L_D+L_{id\_k})\right).$$

(2) 重复包共有 $Q\alpha$%个,这部分包只传输 1 跳,其耗能为

$$Q\alpha\%\,(\varepsilon_{amp}\times(L_D+L_{id\_k})\times d_1^2+E_{elec}\times(L_D+L_{id\_k})).$$

(3) 虚假数据包共有 $Q\beta$%个,由定理 2 可知其耗能为:

$$Q\beta\%\left(\sum_{i=1}^{\frac{mn\left(1-\left(1-\frac{\theta(t-T_c)}{mn}\right)^H\right)}{\theta(t-T_c)}}\varepsilon_{amp}\times(L_D+L_{id\_k})\times d_i^2+\frac{mn\left(1-\left(1-\frac{\theta(t-T_c)}{mn}\right)^H\right)}{\theta(t-T_c)}\times E_{elec}\times(L_D+L_{id\_k})\right),$$

$$Long\_Path=\frac{mn(1-(1-\frac{\theta(t-T_c)}{mn})^H)}{\theta(t-T_c)}.$$

(4) *MNCW* 算法在选取证人、生成和嵌入水印等的预处理的能耗记为 $E_{pretreat}$.

累加得: $E^*_{\text{com}}=Q(1-\alpha\%-\beta\%)\left(\sum_{i=1}^{H}\varepsilon_{amp}\times(L_D+L_{id\_k})\times d_i^2+H\times E_{elec}\times(L_D+L_{id\_k})\right)+$

$$Q\alpha\%\,(\varepsilon_{amp}\times(L_D+L_{id\_k})\times d_1^2+E_{elec}\times(L_D+L_{id\_k}))+$$

$$Q\beta\%\left(\sum_{i=1}^{\frac{mn\left(1-\left(1-\frac{\theta(t-T_c)}{mn}\right)^H\right)}{\theta(t-T_c)}}\varepsilon_{amp}\times(L_D+L_{id\_k})\times d_i^2+\frac{mn\left(1-\left(1-\frac{\theta(t-T_c)}{mn}\right)^H\right)}{\theta(t-T_c)}\times E_{elec}\times(L_D+L_{id\_k})\right)+$$

$$E_{pretreat}.$$

式中 L_D 表示数据包的 Bit 数,d_i 表示每跳的距离 L_{id_k},同样表示密钥编号的 Bit 数.$E_{pretreat}$ 为 MNCW 算法在选取证人、生成和嵌入水印等的预处理的能耗.

4 仿真实验

在本小节,对提出的算法进行仿真实验,实验数据取自热带大气海洋项目(TAO,http://www.pmel.noaa.gov/tao/),实验所用数据集是 TAO 在多个地点的不同深度共 96 个点的传感器从 2004 年 1 月 20 日至 2004 年 5 月 26 日每天的 12:00 采集到的海水温度数据,聚合后的数据 D 为 128×96 的二维数组,实验中我们首先利用算法 1 取 $\xi=16, q=0.05$ 将当前的时间戳和节点的 ID 嵌入到 $LL3$,用于数据内容认证的水印是由数据 D 的三级小波变换得到的小波系数 $LL3$ 嵌入鲁棒性水印后,通过单向函数[4]变换得到的,每个水印共有 100 个 Bit,实验过程中水印仅嵌在数据 D 的小波变换的 10%系数中,故向量 Y 的长度为 128×96×0.1≈1228,子向量 Y_i 的长度 P=12,我们在选择向量 Y 时使用级数较高的小波系数(如: HL_3, LH_3, HH_3, LL_2...).(文中未见图 2-4?)

第 1 个实验的目的是验证水印对原始数据的影响.图 5 给出了聚合数据 D 的三维曲线图,图 6 给出了按 MNCW 算法嵌入 10 个水印之后的聚合数据 D'的三维曲线图,直接从图很难看出两个数据之间的差异,这表明本文算法所嵌入的协作水印具有较好的隐秘性,能有效地迷惑攻击者.图 7 描述了数据 D'和 D 之间的实际差异,

它们之间的差异的绝对值最大不超过 1.5,然而无线传感器网络所采集的数据具有较高的冗余度,这点差异并不会影响用户的决策.故本文的方案是切实可行的.

为了更好地描绘所嵌入水印对原始聚合数据的影响,本文用峰值信噪比(PSNR)和信噪比(SNR)来量化这种影响,表 1 给出了按 MNCW 算法分别嵌入 3~10 个水印的峰值信噪比和信噪比,由表 1 可知,随着水印个数的增加, 峰值信噪比和信噪比呈下降趋势,这表明随着嵌入水印数量的增加对原始聚合数据的影响也增大.推论 1 和定理 3 表明,当水印数量增多时算法在过滤虚假数据的能力和耗能方面均有所改善,但实验结果表明当水印数量增大时对载体数据的影响也变大. 因此算法的性能和采集数据的精度是一个折衷的问题.

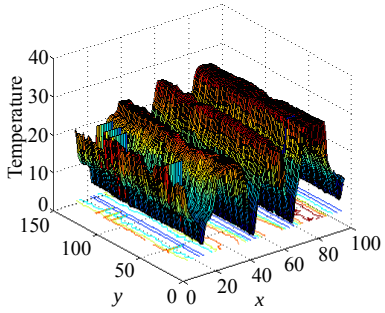

Fig.5 Original aggregation data

图 5 原始的聚合数据

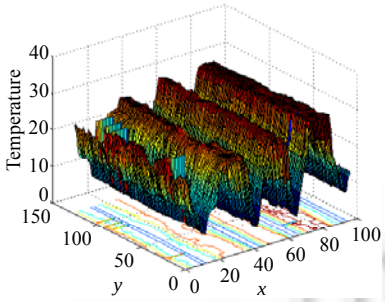

Fig.6 MNCW-Watermarked aggregation data

图 6 按 MNCW 算法嵌入水印之后的数据

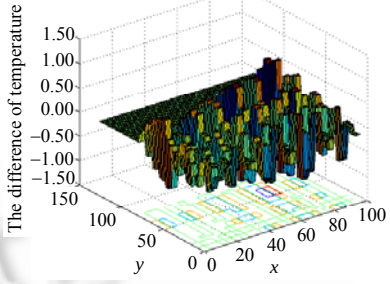

Fig.7 Difference between original data and watermarked data

图 7 图 6 和图 5 的数据差

Table 1 按 MNCW 算法分别嵌入 3~10 个水印的峰值信噪比和信噪比

表 1 按 MNCW 算法分别嵌入 3~10 个水印的峰值信噪比和信噪比

水印个数	3	4	5	6	7	8	9	10
PSNR	45.392 2	43.919 6	43.082 5	42.243 4	41.426 8	40.645 0	39.903 0	39.201 7
SNR	42.157 5	40.684 9	39.847 7	39.008 7	38.192 1	37.410 3	36.668 3	35.967 0

第 2 个实验主要是验证水印本身的鲁棒性和对攻击的敏感性.为了实现数据认证的目的,算法在数据中嵌入 t 个半脆弱性水印,因此希望这 t 个水印对恶意篡改数据和伪造数据具有较高的敏感性,对正常的误差或噪声具有一定的鲁棒性.传感器网络往往部署在复杂的环境,节点之间通过无线通信,节点间的数据传输不可避免地存在误差和噪声干扰,因此认证算法必须要考虑这些问题.

图 8 描绘了当嵌有水印的数据受到不同强度的高斯噪声干扰后,用 MNCW 算法嵌入的多个协作水印的平均受损情况.实验结果表明,本文所提出的算法能承受一定的噪声干扰,当噪声均方差达到 1.5 时,多个协作水印的正确率平均能达到 50%.图 9 描绘了多个协作水印对恶意篡改数据的敏感性,图 9 表明当数据受到恶意篡改时,嵌入在数据内的水印信息同样受到严重的破坏,这说明算法对恶意篡改数据是脆弱的.图 8 纵坐标表示,在嵌有水印的数据包受到不同强度的高斯噪声干扰的情况下,本文的算法仍然将数据包认为是正确的数据包的比率;图 9 纵坐标表示,在嵌有水印的数据包受到不同强度的恶意篡改的情况下,本文的算法仍然将数据包认为是正确的数据包的比率.

第 3 个实验主要是将本文的算法和文献[5−7]进行比较(其余的文献和本文的侧重点不一样),比较的内容主要是如下 3 个方面:首先是能耗;其次是在不改变路由的情况下,算法过滤虚假数据包和重复数据包的概率;最后是在噪声环境和有损压缩的条件下,本文算法和文献[5−7]的虚警率的对比,所谓虚警率是指真实数据包遭受噪声干扰,或受到一定的损失,但数据包本身并没有遭到攻击者的恶意篡改却被算法视为虚假数据并过滤的概率.

图 10 描绘了本文的算法和文献[5−7]算法在能耗上面的差异,实验过程中各参数的值设为 $n=20$, $m=1000$, θ=200,t=10,d_i=40,L_D=1024Bit,L_{id_k}=16Bit,并假定攻击已经俘获一个密钥,当簇头节点发送 100 个数据包(内有 20%的重复包和 50%的虚假数据包)给 sink 节点,图中 E1 表示不采用任何过滤算法时的网络能耗,E2 表示采用

本文的 MNCW 算法时的网络能耗.实验结果表明本文的算法能较文献[5−7]的算法更节能,其主要原因是,文献[5−7]的算法为了实现对数据内容认证的目的,在数据包的后面附加的 t 个 MAC,从而增大耗能,而本文的算法在数据包中嵌入水印来进行认证,代价相对较小.

图 11 描绘了当簇头节点发送 100 个数据包(内有 20%的重复包和 80%的虚假数据包)给 sink 节点,本文的算法和文献[5−7]的算法在过滤虚假数据包和重复包的过滤能力的差异,实验结果表明,本文的算法在过滤概率方面要优于文献[5−7],其主要原因是,本文的算法能识别重复数据包而 SEF 算法不能识别.

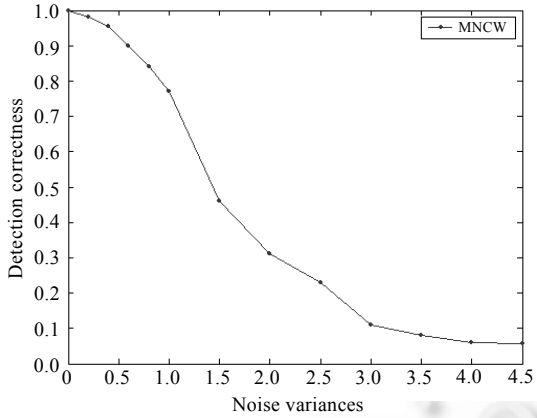

Fig.8 Effects of different intensities noise on multiple collaboration-based watermarks

图 8 不同强度的噪声对多个协作水印的影响

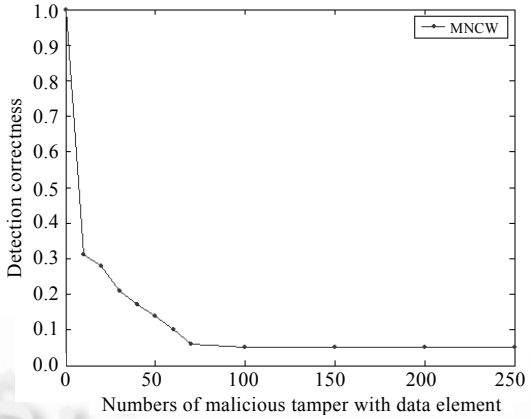

Fig.9 Aensibility on watermark for malevolence hamper data

图 9 水印对恶意篡改数据的敏感性

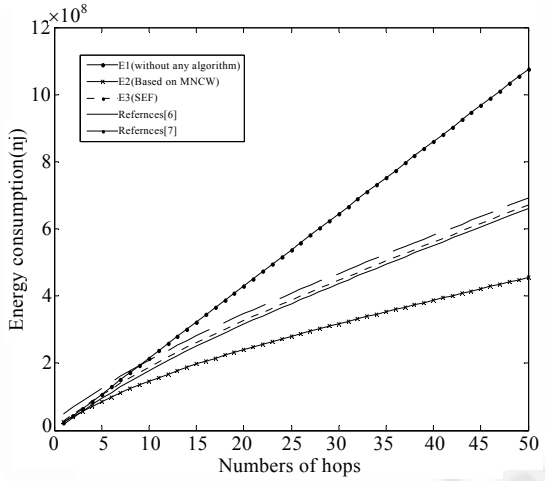

Fig.10 Comparison of energy consumpion between base on MNCW and references [5−7]

图 10 本文的算法和文献[5−7]的能耗对比

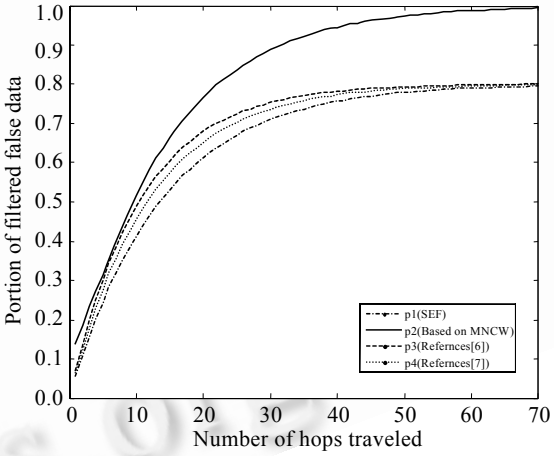

Fig.11 Comparison of filtration capacity for false data between base on MNCW and references [5−7]

图 11 本文的算法和文献[5−7]过滤虚假数据能力的对比

图 12 描绘了在噪声环境下本文的算法和文献[5−7]的算法在虚警率方面的差异,由于本文采用半脆弱水印对数据内容进行认证,半脆弱水印是将数据的三级小波变换的低频系数利用单向函数计算出来,由于数据的三级小波变换的低频系数本身对噪声具有较好的鲁棒性,因此算法具有较好的抗噪声干扰能力,而文献[5−7]的算法在数据包后附加 t 个 MAC,MAC 是利用数据包通过 MAC 码生成算法 MD5 直接计算的,由于 MD5 对参数较为敏感,当数据包被噪声干扰时,计算出来的 MAC 和原始数据的 MAC 有较大的差异,从而出现较高的虚警率.

图 13 描绘了在有损压缩的条件下本文的算法和文献[5−7]的算法在虚警率方面的差异,由于本文将水印三

级小波系数中,而三级小波系本身对有损压缩具有较好的鲁棒性,故算法能抵抗一定品质因子的有损压缩.文献[5−7]的算法是直接在数据包后附加 t 个 MAC,有损压缩将直接损害附加的 MAC 信息,从而出现较高的虚警率.

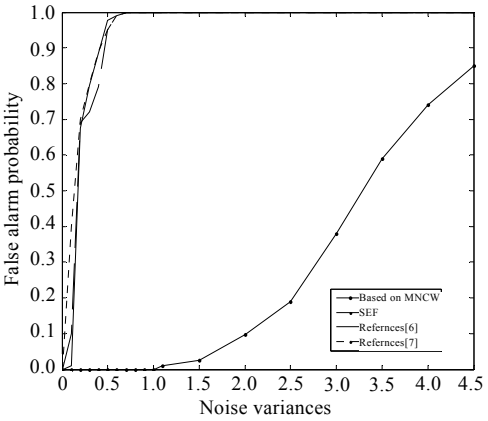

Fig.12 Comparison of false alarm probability under noise environment between base on MNCW and references [5−7]
图 12 在噪声环境下本文的算法和文献[5−7]的虚警率对比

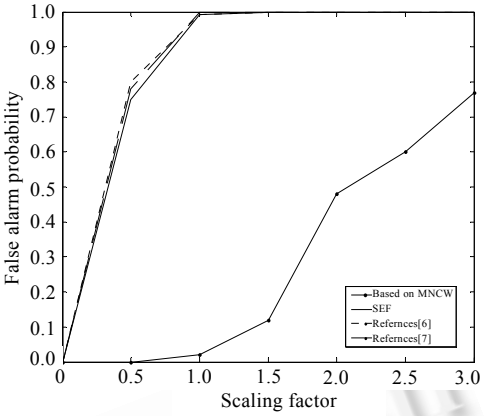

Fig.13 Comparison of false alarm probability under lossy compression condition between base on MNCW and references [5−7]
图 13 在有损压缩条件下本文的算法和文献[5−7]的虚警率对比

5 结 论

如何有效识别和过滤虚假数据包是无线传感器网络中一个具有挑战性的难题,已有的方法过于依赖附加的 MAC,而 MAC 的生成是采用具有高度敏感的 MD5 算法,无线传感器网络通信不可靠、易受干扰等,为了节约能量通常在网内采用有损压缩等,这些特征对 MAC 是有损的,因此这些方法难以付之实用.为此我们提出一种中基于协作水印的虚假数据识别和过滤算法,算法在每个数据包中嵌入两类水印:一类是鲁棒性水印,用于对发送者的身份和数据的新鲜性进行认证;另一类是由 t 个证人节点协作生成、嵌入的半脆弱水印,用于对数据内容进行认证.从而达到对数据包的发送者身份、数据的新鲜性与发送数据内容多重认证的目的.算法保证了多个水印之间互不影响;算法允许网络中的单个节点独立的提取水印,验证数据包的正确性,却不允许伪造或修改水印.理论分析和仿真结果表明本文的算法能在数据转发过程中过滤绝大多数的虚假数据包和重复数据包,节约有限的网络资源,并能对一定程度噪声干扰、有损压缩等具有较好的鲁棒性.本文提出的算法主要用于识别和过滤虚假数据及重复包,如何抵抗其他攻击(如:拒绝服务等)将是我们下一步的工作.

References:

[1] Cui L, Ju HL, Miao Y, Li TP, Liu W, Zhao Z. Overview of wireless sensor networks. Journal of Computer Research and Development, 2005,42(1):163−174 (in Chinese with English abstract).

[2] Li JZ, Li JB, Shi SF. Concepts, issues and advance of sensor networks and data management of sensor networks. Journal of Software, 2003,14(10):1717−172 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/1717.htm

[3] Li P, Lin YP, Zeng WN. Search on security in sensor networks. Journal of Software, 2006,17(12):2577−2588 (in English with Chinese). http://www.jos.org.cn/1000-9825/17/2577.htm

[4] Ye F, Luo HY, Lu SW, Zhang LX. Statistical en-route filtering of injected false data in sensor networks. In: Proc. of the 23rd Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM 2004). Hong Kong: IEEE Press, 2004. 2446−2457.

[5] Ye F, Luo HY, Lu SW. Statistical en-route filtering of injected false data in sensor networks. IEEE Journal on Selected Areas in Communication, 2005,23(4):839−850.

[6] Zhu SC, Setia SJ, Jajodia S, Ning P. An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks. In: Proc. of the IEEE Symp. on Security and Privacy. Berkeley: IEEE Press, 2004. 259−271.

[7] Yang H, Lu SW. Commutative cipher based en-route filtering in wireless sensor networks. In: Proc. of the IEEE 60th Vehicular Technology Conf. Los Angeles: IEEE Press, 2004. 1223−1227.

[8] Yang CG, Xiao J. Location-Based pairwise key establishment and data authentication for wireless sensor networks. In: Proc. of the 2006 IEEE Workshop on Information Assurance. New York: IEEE Press, 2006. 247−252.

[9] Ma M. Resilience of sink filtering scheme in wireless sensor networks. Computer Communications, 2006,30(1):55−65.

[10] Zhou L, Ravishankar, CV. A fault localized scheme for false report filtering in sensor networks. In: Proc. of the 2nd Int'l Conf. on Pervasive Services (ICPS 2005). Santorini: IEEE Press, 2005. 59−68.

[11] Hu YP, Lin YP, Yu JP, Zeng WN. A robust authentication scheme for filtering of injected false data in dynamic wireless sensor networks. Chinese Journal of Electronics (CJE), 2007,(3):443−448 (in Chinese with English abstract).

[12] Yu Z, Guan Y. A dynamic en-route scheme for filtering false data injection in wireless sensor networks. In: Proc. of the 25th IEEE Int'l Conf. on Computer Communications (INFOCOM 2006). Barcelona: IEEE Press, 2006. 1−12.

[13] Zhang YT, Yang J, Vu HT. The interleaved authentication for filtering false reports in multipath routing based sensor networks. In: Proc. of the 20th Int'l Parallel and Distributed Processing Symp. (IPDPS 2006). Rhodes Island: IEEE Press, 2006.112−121.

[14] Li F, Wu J. A probabilistic voting-based filtering scheme. In: Proc. of the 2006 Int'l Wireless Communications and Mobile Computing Conf. British Columbia: IEEE Press, 2006. 27−32.

[15] Wong Peter HW, Au Oscar C, Yeung YM. A novel blind multiple watermarking technique for images. IEEE Trans. on Circuits and Systems for Video Technology, 2003,8(13):813−830.

[16] Feng J, Potkonjak M. Real-Time watermarking techniques for sensor networks. In: Proc. of the SPIE—The Int'l Society for Optical Engineering. Santa Clara: SPIE Press, 2003. 391−402.

[17] Zhou SW, Lin YP, Wang JL, Zhang JM, Ouyang JC. Compressing spatial and temporal correlated data in wireless sensor networks based on ring topology. In: Proc. of the 7th Int'l Conf. on Web-Age Information Management (WAIM 2006). Hong Kong: Springer-Verlag, 2006. 337−348.

[18] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient communication protocol for wireless microsensor networks. In: Proc. of the Hawaii Int'l Conf. on System Sciences. Springer-Verlag, 2003. 181−191.

[19] Yu GJ, Lu CS, Liao HYM. Mean-Quantization-based fragile watermarking for image authentication. Optical Engineering, 2001,40(7):1396−1408.

附中文参考文献:

[1] 崔莉,鞠海玲,苗勇,李天璞,赵泽.无线传感器网络研究进展.计算机研究与发展,2005,42(1):163−174.

[2] 李建中,李金宝,石胜飞.传感器网络及其数据管理的概念、问题与进展.软件学报,2003,14(10):1717−1727. http://www.jos.org.cn/1000-9825/14/1717.htm

易叶青(1977－),男,湖南邵阳人,博士生,讲师,主要研究领域为无线传感器网络中的安全机制,机器学习.

林亚平(1955－),男,博士,教授,博士生导师,主要研究领域为计算机网络,机器学习.

李小龙(1981－),男,博士生,主要研究领域为传感器网络覆盖控制,定位技术

羊四清(1966－),男,博士,副教授,主要研究领域为传感器网络数据处理.

尤志强(1972－),男,博士,副教授,主要研究领域为嵌入式系统.